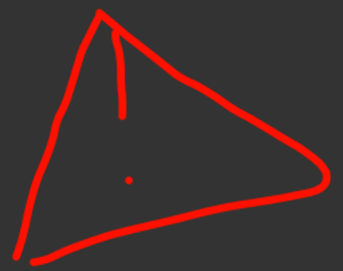
Python Machine Learning Labs

Good morning!

We start at 10:00



Machine learning Pipeline

Problem Statement → EDA → FE → Modeling → Fine-Tuning

Pclass Encoding

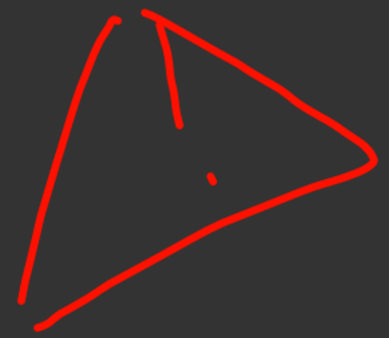
Pclass
1
2
3

Pclass	1	2	3
1	1	0	0
2	0	1	0
3	0	0	1

Dataframe Concatenation

Row-wise

Column-wise



Dataset Splitting (1)

Train

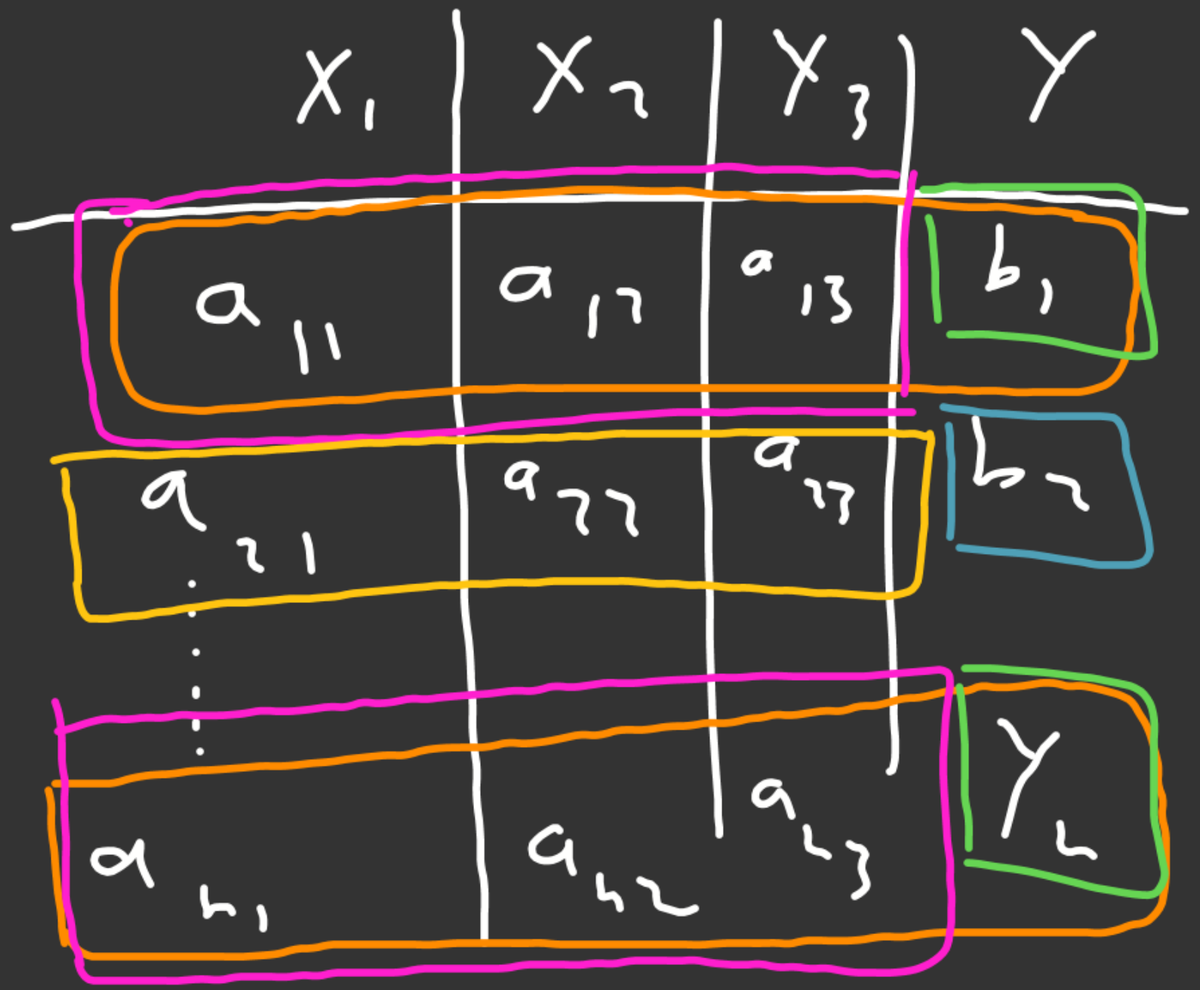
75 - 80 %

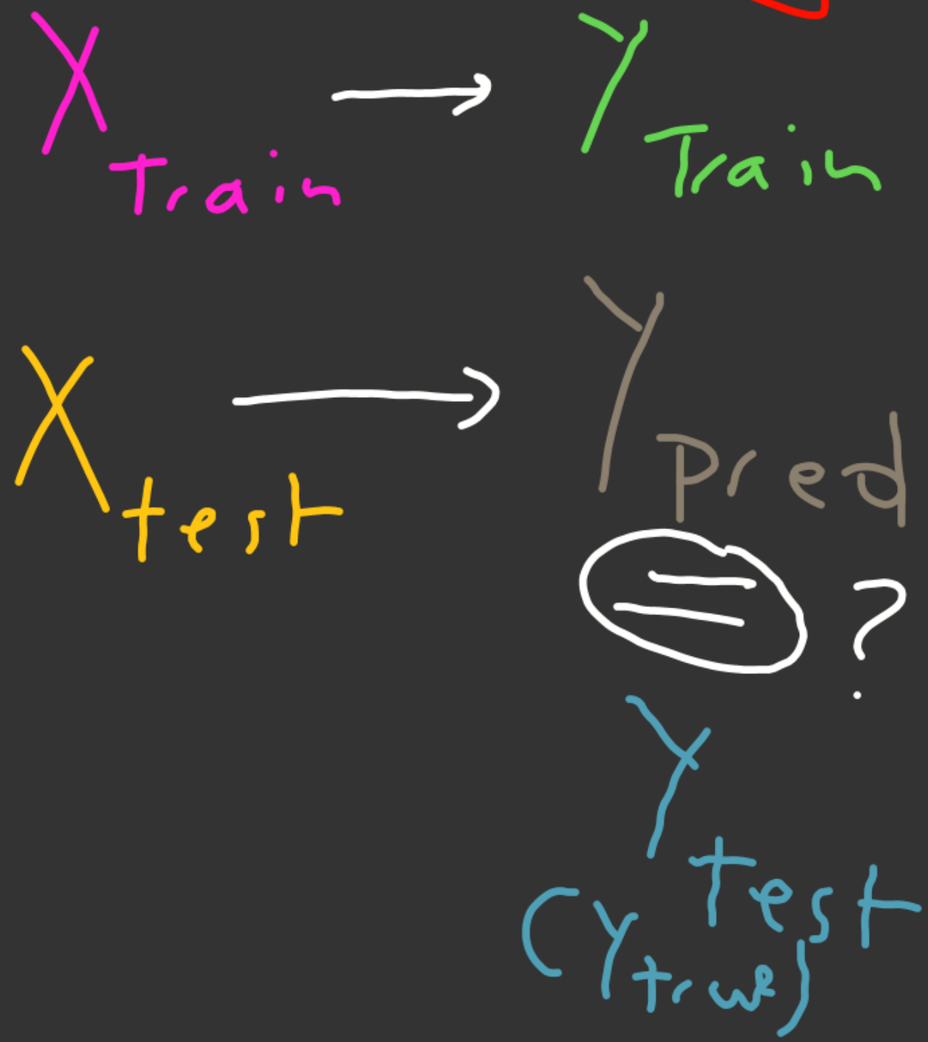
Test

20 - 25 %

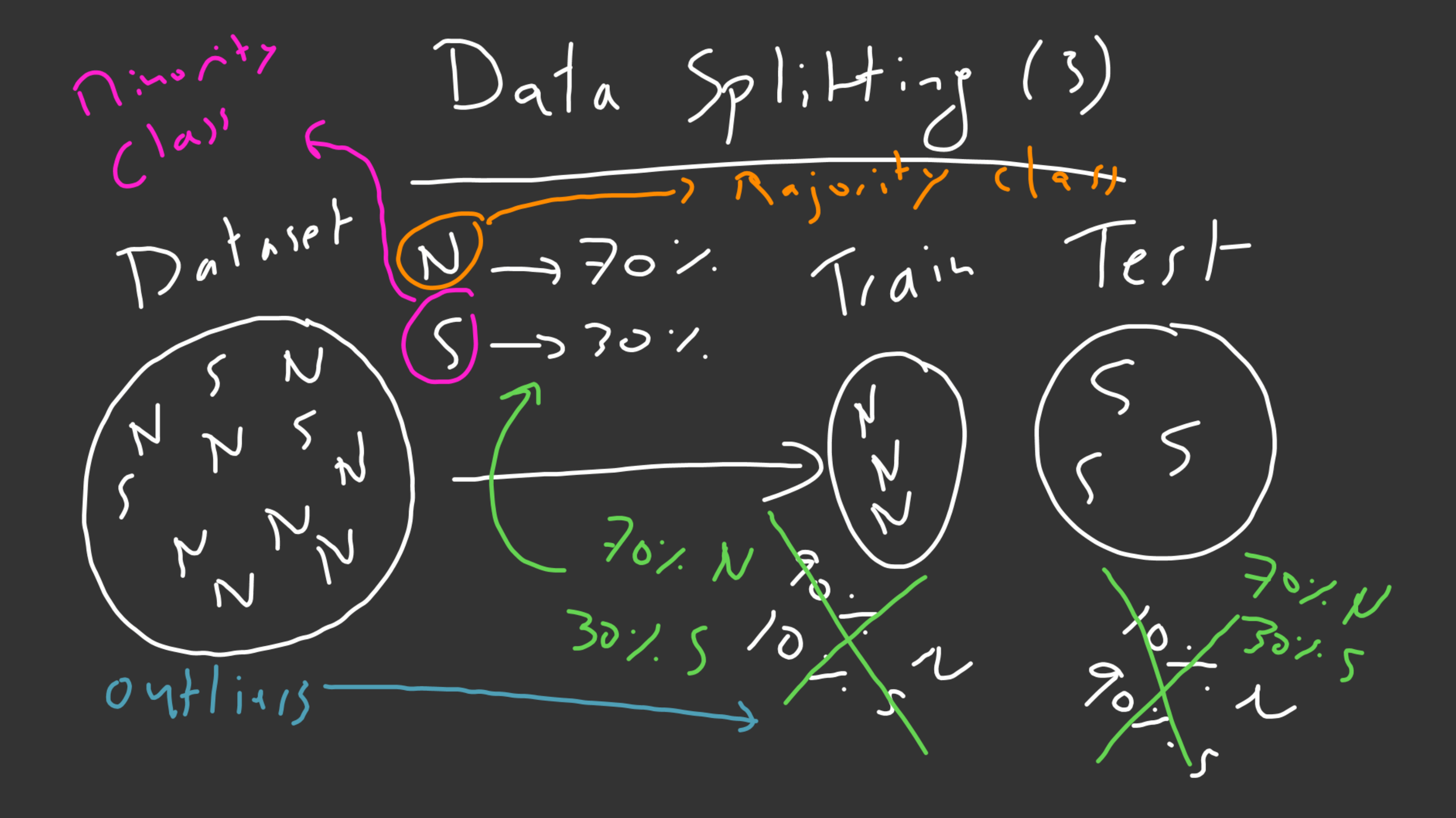
- Train set
- Test set

Data Splitting (2)





Data Splitting (3)



Break

Back at 11:45

Different types of learning...

Supervised

Labeled data

e.g.

Unsupervised

Unlabeled data

e.g.

Semi (self) - supervised

Small part of data labeled + a big part unlabeled

e.g.

Document classification

Reinforcement learning

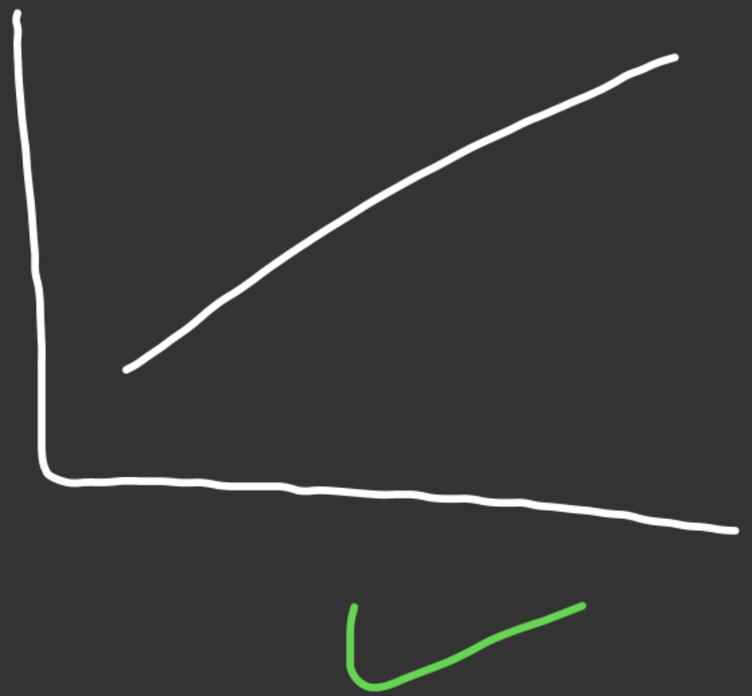
Model learns from environment (data)

e.g.

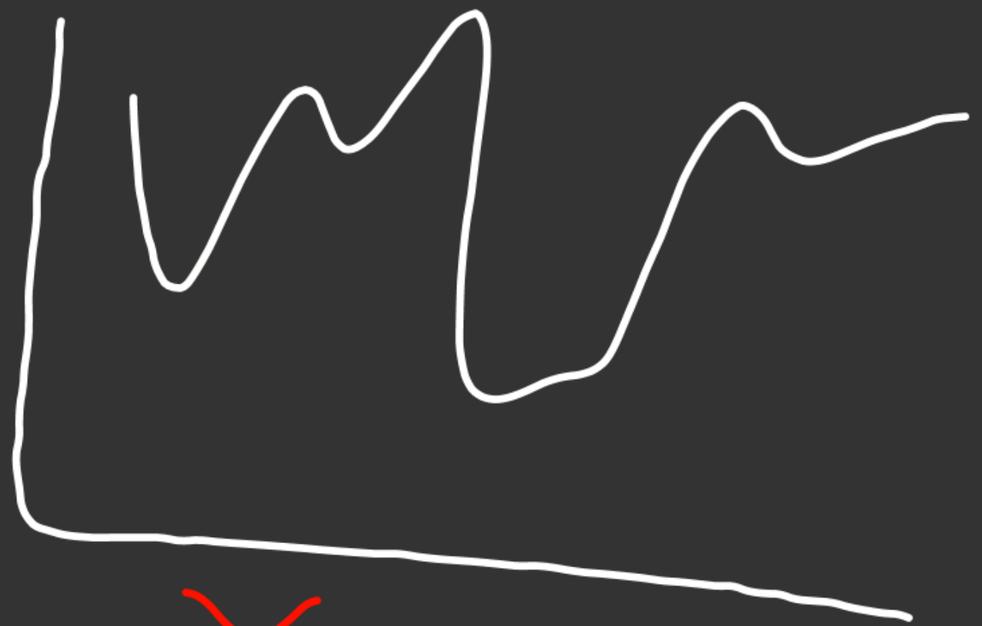
Chess (Game)

Data Distributions

Linear



Non-Linear



Linear Models

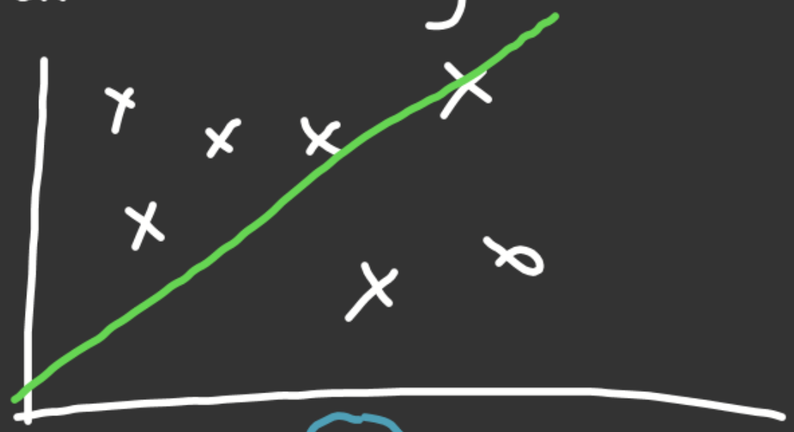
Classification

Logistic Regression

$$Y = \frac{1}{1 + e^{-X}}$$

Regression

Linear Regression



$$Y = aX + b$$

Y: target

X: Features

Lunch

Back at 14:00

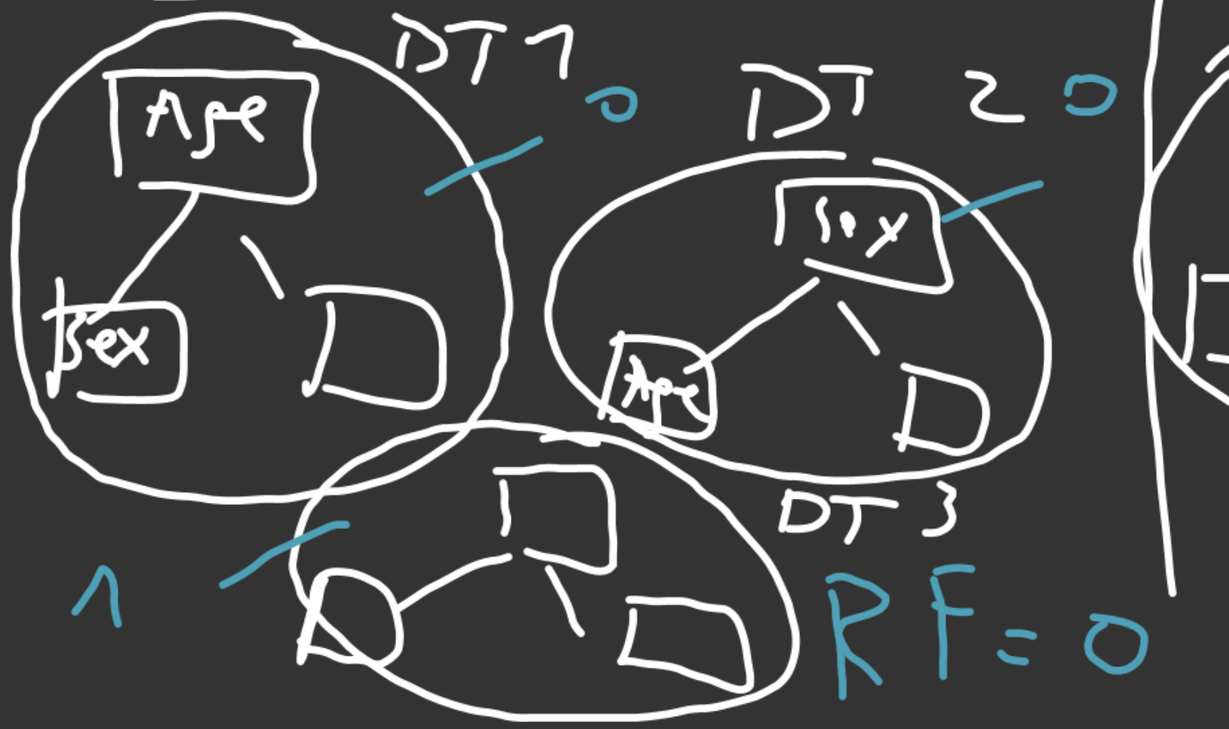
Non - Linear Models (1)

Decision Trees

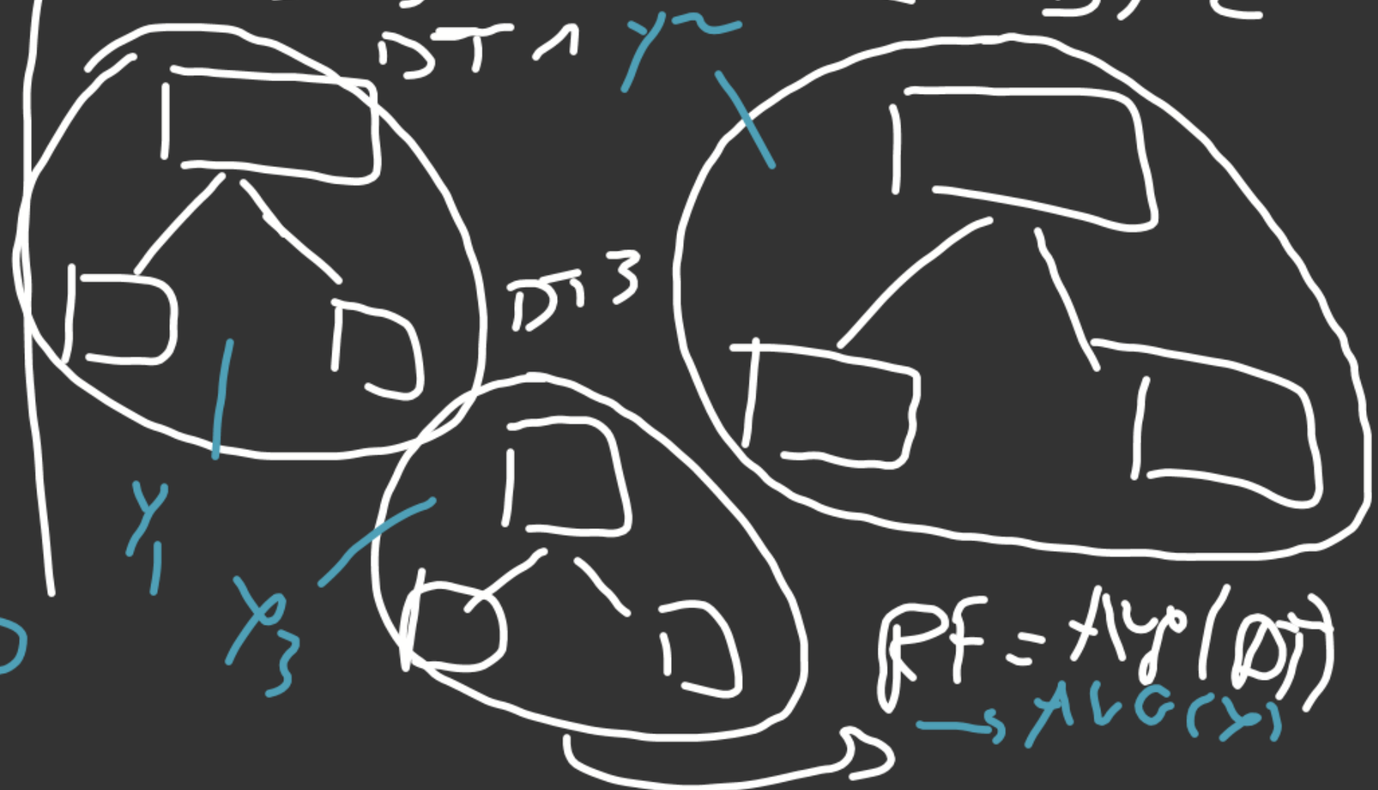
Non-Linear Models (2)

Random Forest

Classification

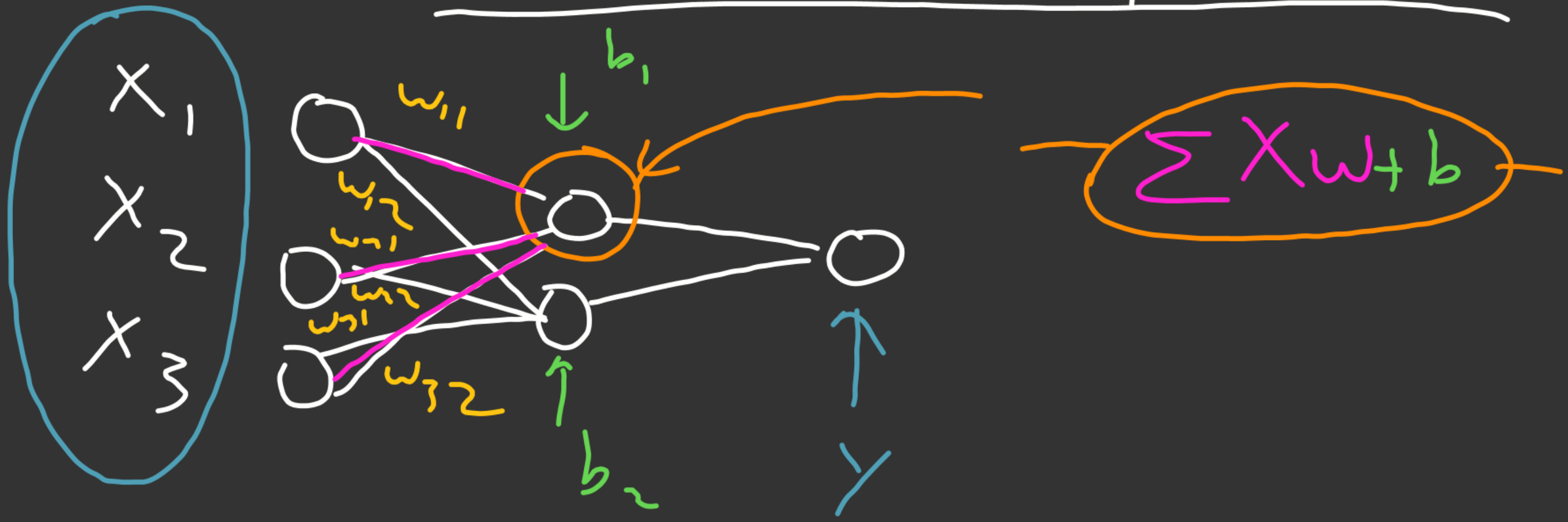


Regression



Non-Linear Models (3)

Multi-layer Perceptron (MLP)



Random State in ML Models

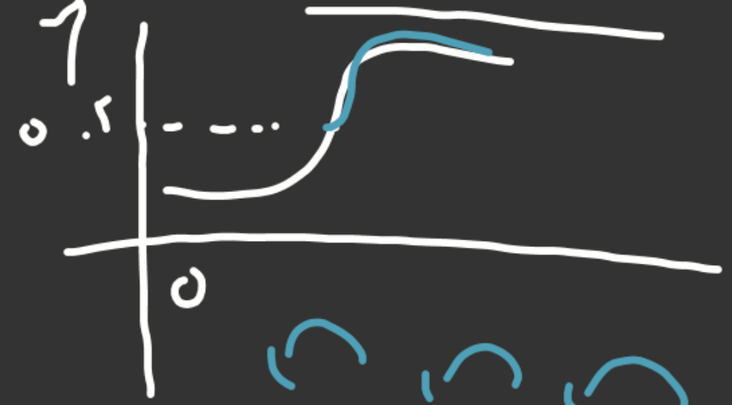
Logistic Regression

Case 1 (Run 1)

$Y = 0.499$

$< 0.5 \rightarrow$ class 0

Case 2 (Run 2)



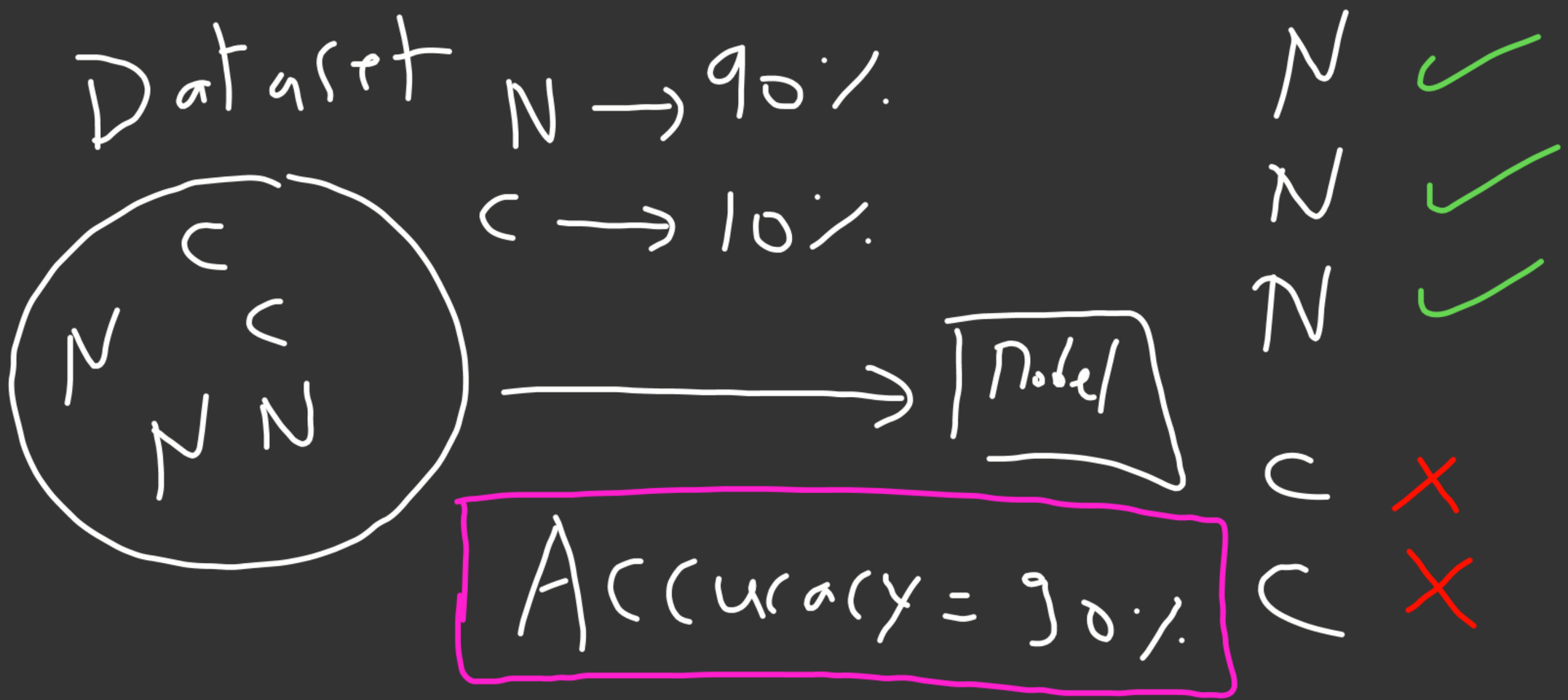
$Y = 0.499$

$= 0.5 \rightarrow$ class 1

Break

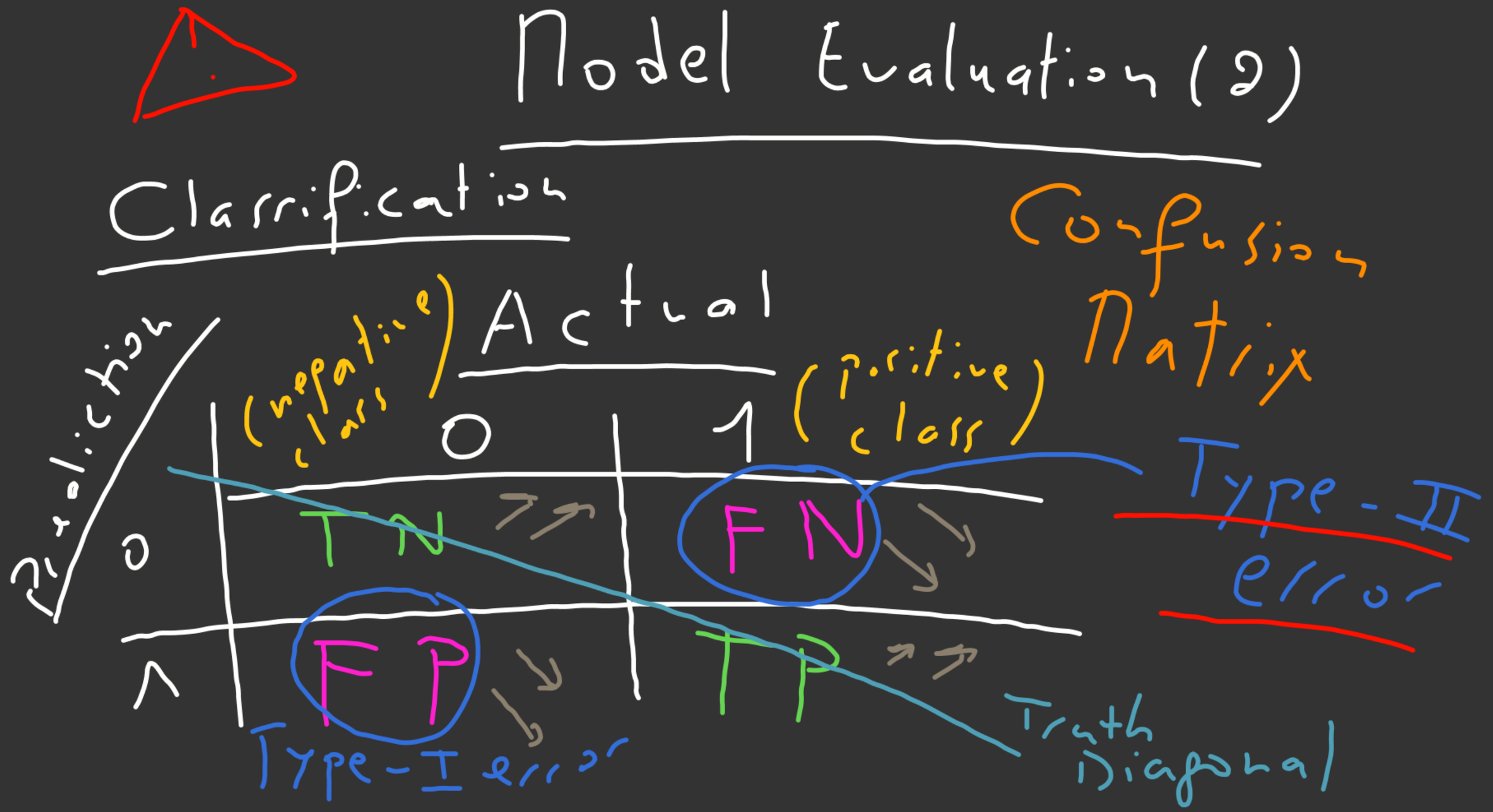
Back at 15:05

Model Evaluation (1)



Model Evaluation (2)

Classification



Model Evaluation (3)

Precision

$$Pr = \frac{TP}{TP + FP}$$

FP: Type-I error

Recall

$$Re = \frac{TP}{TP + FN}$$

FN: Type-II error

Model Evaluation (4)

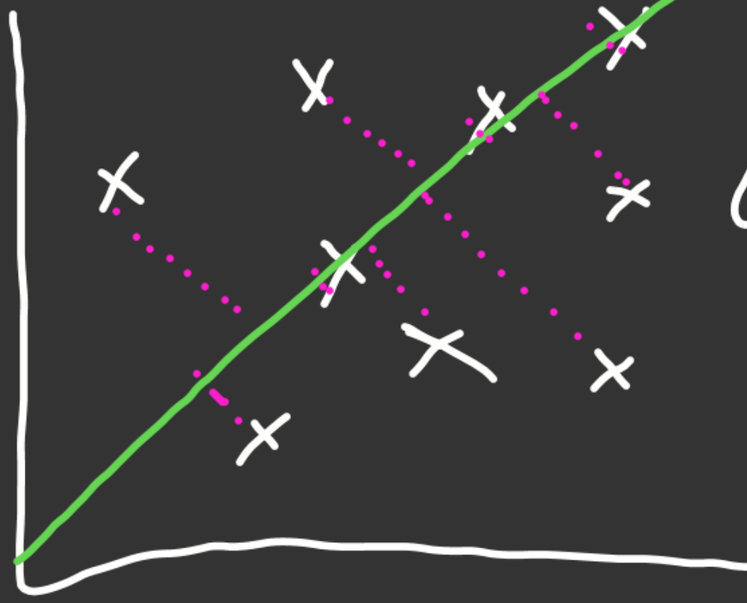
FP FN

$$F1\text{-}score = 2 \cdot \frac{Pr \times Re}{Pr + Re}$$

(classification)

Model Evaluation (5)

Regression:



MSE:

$$\frac{1}{n}\sum (y_T - y_p)^2$$

$$RMSE = \frac{\sqrt{\sum (y_T - y_p)}}{n}$$

Modeling phase

- Select models
- Evaluate
- Compare
- Interpret (Analyze)

Fine-Tuning phase

- Changing model ←
- Change Features
- Change data